DOCUMENT RESUME

ED 379 293 TM 022 638

AUTHOR Lang, William Steve; And Others
TITLE Rasch Model Applications To Determine the Equivalence of a Readiness Test in Two Languages.
PUB DATE Apr 93
NOTE 21p.; Paper presented at the International Objective Measurement Workshop (Atlanta, GA, April 10, 1993). Table 4 contains very small, filled print.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Bias; Blacks; Computer Software; Cultural Awareness; *Cultural Differences; English; Goodness of Fit; Hispanic Americans; *Item Response Theory; Multilingual Materials; *Preschool Children; Preschool Education; School Readiness; *School Readiness Tests; Sex Differences; Spanish; Statistical Analysis; Test Items; *Test Validity; Whites
IDENTIFIERS *Lollipop Test; *Rasch Model

ABSTRACT

The Lollipop Test (La Prueba Lollipop) is a bilingual preschool readiness test (in both English and Spanish) that has been the subject of a number of studies to assess validity and detect cultural bias. Such studies have not dealt with item analysis as a way to measure cultural fairness. The Rasch model was used in a study of the Lollipop test which considered its cultural bias, the usefulness of the Rasch model, and the utility of the new software, IPARM. Subjects were 61 4- and 5-year-olds in Georgia and Florida (25 white, 24 black, and 12 Hispanic Americans). Results do not suggest gender or cultural bias for the test as a whole. The potentially biased items favor female or Hispanic students. The Rasch approach was useful, and the software was easy to use, presenting results on item functioning in easily understood format. Five tables present study data. (Contains 10 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 379 293

Rasch Model Applications to Determine the Equivalence of a Readiness Test in Two Languages

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
WILLIAM STEVE LANG

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

William Steve Lang
University of South Florida at St. Petersburg

Alex L. Chew
Georgia Southern University

Cecilia Vargas
University of South Florida at St. Petersburg

Presented at the International Objective Measurement Workshop
Emory University
Atlanta, Georgia
April 10, 1993

The authors would like to express appreciation to Richard Smith for the patient answering of questions, allowing visitors to his classes, and his interpretation of PIE.

TM022638

Rasch Model Applications to Determine the Equivalence of a Readiness Test in Two Languages

Bilingualism has been acknowledged to be a complex problem in psychoeducational assessment for over half a century (Arsenian, 1937). Several studies have commented on the serious drawbacks of using standardized, commercially developed, English language assessment instruments for bilingual students or as translations without intensive comparative research (Figueroa, 1983, Mardell-Czudnowski, 1987).

The *Lollipop Test (La Prueba Lollipop)* is a preschool readiness test in both English and Spanish which has been the subject of a number of studies to assess validity and detect cultural bias using correlation, regression and discriminant analysis statistics (Chew & Lang, in press, Lang, Chew & Shomber, 1991, Chew & Lang, 1990). Unfortunately, these studies have focused on subtest or total scores and have not dealt with item analysis as a way to measure cultural fairness or bias. The primary problem with classical item analyses is sample dependency. Unless the same person can take both language forms of a test, the items cannot be easily compared for parallel functioning in classical item statistics. Preschoolers are rarely proficient in one language, much less two languages, so that these comparisons can be made.

Rasch model statistics are useful here for two reasons. One is the sample independence of the item analyses. The other is the recently developed applications of between fit statistics in a useful computer application (Smith, 1991). Rather than treat this as a test equating exercise with common-item or common-person calibration, the scores here are to be pooled in a single calibration while culture is treated as a demographic (like gender or race) in a bias detection approach. IPARM (Item and Person Analysis with the Rasch Model, 1992) offers new capabilities and graphic solutions to the task of quality control of person and item measures. For a fuller discussion of detecting item bias using the Rasch model, see Smith (1992).

Of particular interest in this study was the use of between fit statistics for the detection of bias (Smith,1991). As a powerful statistic for identifying measurement disturbance, between fit was most useful. Finding a qualifying sample of Hispanic preschoolers is relatively difficult and power with a small number of subjects was considered more important than the possibility of Type I error. Naturally, a signal that bias was present in test items, even by chance, would lead to conservative interpretation of scores and judgmental examination of the items instead of misplaced trust in the test results. In other words, Type I errors would not lead to concluding a test was culturally fair, so Type II errors were more practically to be avoided.

The research design was intended to answer several questions. First, does the Spanish translation of the test perform the same as the English form on an item by item comparison? Second, do the Rasch model results parallel the previous classical studies? Are there any suggestions for using fit statistics and Rasch model analyses for the particular use of cross-lingual test development which followed from the experience? What is the utility of the relatively new software, IPARM?

METHOD

Subjects

The subjects were 61 four and five year old preschoolers from public preschools and kindergartens in Georgia and Florida. A total of 7 schools were part of the data collection. The sample consisted of 25 white, 24 black and 12 Hispanic children. The original sample was also split approximately in two by gender with 30 male and 31 female participants.

The data were collected in March and April of 1992 and 1993 by examiners trained in the administration of the test. For Hispanic children, the examiner spoke both English and Spanish. As is typical with bilingual children in

the United States, they often spoke Spanish at home and English at school. The examiner informally asked the Hispanic children if Spanish was the language spoken at home and answered any questions regarding the testing permission for Spanish-speaking parents.

Instrument

La Prueba Lollipop (Chew, 1989) is an individually administered, criterion-referenced screening measure of school readiness consisting of the following four subtests: (1) Identification of Colors and Shapes; (2) Picture Description, Position, and Spatial Recognition; (3) Identification of Numbers and Counting; and (4) Identification of Letters and Writing. The test items are individually administered orally with a total score range of 0 to 69. Preliminary investigation of the Spanish edition of *La Prueba Lollipop* (Lang, Chew, and Schomber, 1992) found no evidence of systematic bias for bilingual (Spanish) groups. That study compared three groups, notably alike in major demographic characteristics and found evidence that the Spanish and English versions of the *Lollipop Test* performed similarly with students of comparative socioeconomic status. Conclusions drawn from that effort suggested construction of the test, and the measurement of school readiness has not been confounded with culturally loaded items, a problem often seen in test translations. The English version of the test has been found to be relatively independent of socioeconomic variables and requires approximately 15 to 20 minutes to administer and score (Chew and Morris, 1987). The concurrent validity (Chew and Morris, 1987) and the predictive validity (Chew and Lang, 1990) is well-documented in the literature.

Procedure

All children were identified by principals and school district administrators as eligible for this study. The requirements were simply that the students fall

within the age-range of the instrument, and that there was no objection from a parent or school official to testing the child. Each child was tested individually according to the standardized directions.

Statistical Analysis

Even though *The Lollipop Test* totals 69 possible points, the value of every item is not always one point in scoring. Some items award two points and some five points to a total. For purposes of analysis, the test items were entered as single point, dichotomous data. The result was that 58 separate items were included.

The data were first calibrated using the Rasch Model and the BIGSTEPS program. This created an initial item difficulty file with the associated fit statistics and item/person maps. The difficulty file was then used in a subsequent analysis using the IPARM software. In the second analysis, item subpopulation analyses were performed for ABILITY (three groups), SEX (male and female), and CULTURE (white, black, and Hispanic). For each of these breakdowns, between fit statistics, distracter analysis, and the predicted/observed proportion were produced. Person analyses were generated for each of the four subtests. For a complete discussion of BIGSTEPS and IPARM, see Smith (1991; 1992).

Several points are in order here. All items were used in the IPARM analysis regardless of the initial item fit statistics generated by BIGSTEPS. All persons were included in the IPARM analysis regardless of the initial person fit statistics generated by BIGSTEPS. The rationale here was to have both the overall misfitting items and those items showing potential bias available simultaneously. Gender was known to be a variable without bias since the test had been examined for this before. It was included as a variable as a comparison to a unknown factor, culture.

This was the first known Rasch analysis of *The Lollipop Test*. As such, the subtest scores were included as an analysis in addition to the test as a whole. This

test has several clearly different types of items related to preschool readiness. For example, recognition of numbers, letters and spatial relations of objects. For this analysis, the test was simply broken down into the subtests. It is of course possible that other items classifications (such as items which require pointing to respond, telling to respond, and drawing to respond) would make sense, but there is no reason at the moment to speculate better than the test author has designated. The fit of items to meet the requirements of the Rasch model was the intent here. For all fit statistics, 2.0 (95% confidence level) was considered the criteria for examination.

RESULTS & DISCUSSION

Initial Calibration Results

The results of BIGSTEPS analysis for *The Lollipop Test* is summarized in Table 1. The item/person map reveals that our sample was particularly strong for the range of the test. All 58 items entered the analysis with 58 persons. Three persons were dropped. Two persons obtained perfect scores while one person was an judged a misfit with a 2.07 infit and a 3.60 outfit. There is the possibility that one person of 61 was simply a Type I error. An examination of that person report revealed a generally weak ability (-1.12) with unusually high scores in one particular subtest (a spontaneous response to a picture). The score seemed to reflect selective knowledge in that area with the obvious likelihood that this preschooler had been exposed to the subtest material in some enriched way (comparatively) or was shy and only decided to respond to the area of questions favored.

Ten of the 58 items (8,10,14,18,19,21,22,23,25,28) had both infit and outfit greater than 2.0. Three of these items were from subtest 1 and were the more difficult of the the shape recognition and drawing tasks. Five items were

from subtest 2 and dealt with the more difficult spatial relations (left, underneath, first, last). Two items were from subtest 3 and involved the recognition of numbers. No items from subtest 4 were revealed as potential problems. Items 8, 18, 19, 21, and 22 were among the most difficult on the test (1.43, 1.16, 2.25, 1.29, and 2.25 respectively). Since the children had opportunity to "guess" by pointing randomly at the stimulus card, it is quite likely the disturbance was due to guessing. Item 14 involved drawing a square which was dependent on motor coordination and scoring effects. Items 21, 23, 25, and 28 were subject to guessing, but not as difficult. Item results are given in Table 2.

Item and Person Analyses

Two of the 58 items (5 & 28) revealed a between fit statistic greater than 2.0 for sex differences. Both show advantage to female students. Item five deals with color recognition (Brown) while item 28 is number recognition ("9"). Even though it is possible to imagine color recognition being gender related, there is no explanation for item 28 except to state again that the item seems to be subject to guessing as revealed in the first analysis. These item profiles are shown in Table 3. The overall gender between fit statistic was .01 indicating virtually no bias. It is possible that one or two items are simply Type I error.

It is interesting to note that item five was spontaneously singled out by our interpreter as a possible problem for Hispanics since the translation to "brown" was not what she suggested and some children gave the answer "chocolate." There was no culture bias revealed for this item as it had a between fit of -.10.

Three items appeared to show culture bias with a between fit greater than 2.0 (9, 19, & 25). Again, two of these (19 & 25) had already been identified as problems in the initial calibration. In all three cases the bias is in favor of Hispanic students with differences between the predicted and observed proportions of .223, .284, and .264 respectively. The item profiles are summarized in Table 4. The overall between fit for culture was .24 and no items

were revealed to be bias where black students were the criteria of 2.0 was met. The worst possible item (11) in this regard had a between fit of 1.84 and the black student difference was -.155.

Sixteen of the 61 persons had an unweighted total fit or a weighted fit or a subtest between fit greater than 2.0. Since the subtests differed in content that might be taught simultaneously at school (such as letters and numbers or shapes and colors), but might not be emphasized with equal experience at home, it is quite likely that these preschoolers are subject to the whims of parental values and the lack of fit is somewhat expected at this early age.

Conclusions

There does not appear to be any gender or culture bias for *The Lollipop Test* as a whole. Even the items which do not meet the between fit criteria of 2.0 are mostly items which are a subset of misfitting items on the whole. It is interesting to note that all of the potentially biased items are in favor of female or Hispanic students. It seems that the author of the test has little to worry about with regard to underscoring traditional minorities.

One hypothesis that was suggested above is that the items showing misfit were a result of the child guessing by randomly pointing to the answer on the stimulus card. Since some figures on a card are already used in earlier items by the time the child gets to the "misfitting" item, the child is guessing among the figures left over. Smith (1991) suggests that the power of the between fit statistic is useful in detecting systematic bias such as cultural or gender while the unweighted total fit statistic is more sensitive to random disturbances such as guessing. In examining the total unweighted fit statistics of the 5 potentially biased items, two (19 and 28) revealed evidence of possible guessing while three did not result in values greater than 2.0. In fact, when one looks at the total unweighted fit for all ten potential misfitting items, the contrast between those with high total unweighted fit and those with high between fit from the item and

person analysis is evident. These figures are summarized in Table 5.

Based on Table 5, one might suggest that items 8, 19, 22, 23, and 28 seem to reflect more random disturbance and race or gender bias might be related to that characteristic, be it guessing or some other factor. On the other hand item 5 seems to reflect systematic gender bias while items 9 and 25 might reflect systematic culture bias. Since 25 is a very difficult item and it shows an unweighted total fit of 1.14, there is a possibility that is random disturbance with a ceiling effect.

Regardless of the final determination of the individual items, it seems appropriate to conclude with a statement about the utility of this type of analysis. Since this is a relatively small sample and a large number of statistics are utilized, the potential for some Type I error is great. Fortunately, that is fine when one considers the result of that error is likely that a test will not be used inappropriately.

One problem is that the small sample used is likely to be a power problem where the analysis is less likely to detect bias. The choice of the between fit statistic was an attempt to offset that possibility, but there is only so much that can be done with subgroups of 12, 24, and 25. Quite likely more items would have been suggested as the sample size increased, but at a certain point the power would have become so great that proportional difference which are practically meaningless become significant. The test examiner must find some balance here.

Finally, the IPARM software is a welcome addition to those who want to quickly and easily get results that send them back to the test with a serious intent to revise and edit. The results reveal many clues to item functioning in easily understood format.

REFERENCES

Arsenian, S. (1937). *Bilingualism and mental development.* New York: Columbia University.

Chew, A.L. (1989). *The Lollipop Test: A Diagnostic Screening Test of School Readiness-Revised.* Atlanta, GA: Humanics Limited.

Chew, A.L. & Morris, J.D. (1987). Investigation of the *Lollipop Test* as a prekindergarten screening instrument. *Educational and Psychological Measurement, 47,* 987-471.

Chew, A. & Lang, W. S. (in press). Concurrent Validation and Regression Line Comparison of the Spanish Edition of the *Lollipop Test (La Pruebe Lollipop)* on a Bilingual Population. *Educational and Psychological Measurement.*

Chew, A. L. & Lang, W. S. (1990). Predicting academic achievement in kindergarten and first grade from prekindergarten scores on the Lollipop Test and Dial. *Educational and Psychological Measurement, 50,* 431-437.

Lang, W.S., Chew, A.L. & Schomber, J. (1992). The comparative use of the Spanish and English editions of the *Lollipop Test* : a cross cultural study. *Journal of Research in Education, 2 (1),* 23-30 .

Figueroa, R. A. (1983). Test bias and Hispanic children. *The Journal of Special Education, 17* (4), 431-440.

Mardell-Czudnowski, C. D. (1987). Screening kindergartners at an international school. *School Psychology International, 8,* 127-132.

Smith, M. R. (1991). *IPARM: Item and person analysis with the Rasch Model.* Chicago: Mesa Press.

Smith, M. R. (1992). *Applications of Rasch Measurement.* Chicago: Mesa Press.

Table 1

Summary of BIGSTEPS Rasch Analysis of *The Lollipop Test*
Map of Persons and Items

SUMMARY OF 58 MEASURED (NON-EXTREME) PERSONS

	SCORE	COUNT	MEASURE	ERROR	MNSQ	INFIT	MNSQ	OUTFIT
MEAN	40.7	57.7	1.71	.49	.99	.0	.96	.2
S.D.	15.0	2.1	1.98	.23	.15	.9	.55	.9

RMSE .54 ADJ.S.D. 1.90 PERSON SEP 3.51 PERSON SEP REL. .92

MAXIMUM EXTREME SCORE: 2 PERSONS
LACKING RESPONSES: 8 PERSONS

SUMMARY OF 58 MEASURED (NON-EXTREME) ITEMS

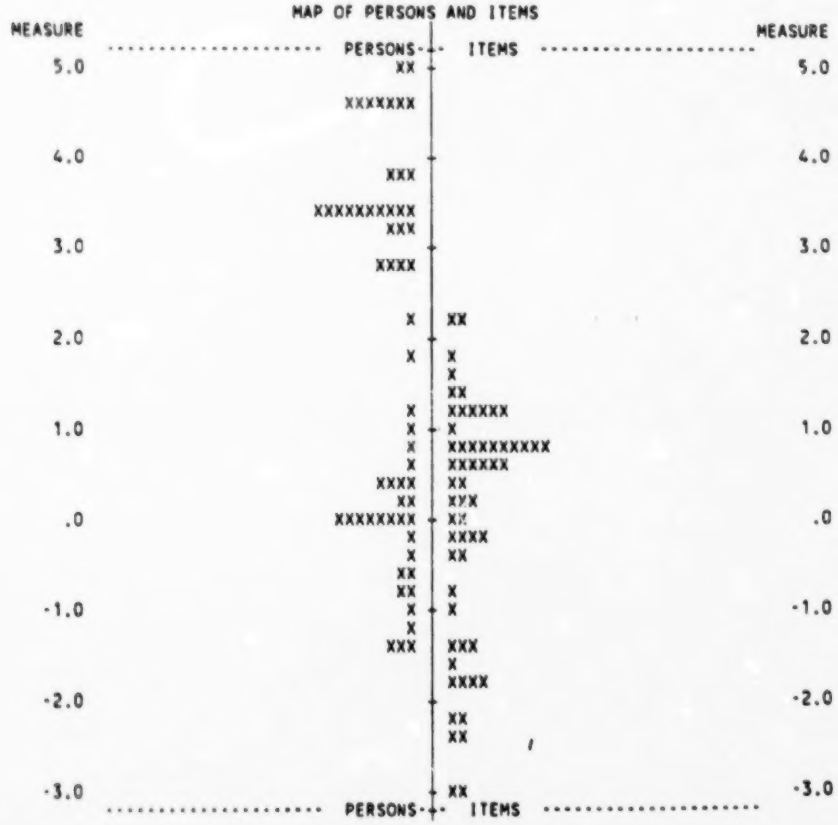
	SCORE	COUNT	MEASURE	ERROR	MNSQ	INFIT	MNSQ	OUTFIT
MEAN	40.7	57.7	.00	.41	.99	-.1	.96	.0
S.D.	8.5	.4	1.32	.09	.32	1.5	.76	1.3

RMSE .42 ADJ.S.D. 1.25 ITEM SEP 3.00 ITEM SEP REL. .90

SUMMARY OF MEASURED STEPS

LABEL	VALUE	COUNT
0	0	988
1	1	2360

MAP OF PERSONS AND ITEMS



BEST COPY AVAILABLE

Table 2

Summary of BIGSTEPS Analysis of *The Lollipop Test*
Item Statistics and Map

ITEMS STATISTICS: MEASURE ORDER

NUM	SCORE	COUNT	MEASURE	ERROR	MNSQ	INFIT ZSTD	MNSQ	OUTFIT ZSTD	PTBIS	NAME
19	25	58	2.25	.37	1.77	2.8	2.56	3.1	.34	10019
22	25	58	2.25	.37	1.59	2.3	2.20	2.5	.42	10022
40	29	58	1.70	.37	1.11	.5	1.19	.6	.68	10040
44	30	57	1.53	.37	.69	-2.6	.60	-2.1	.87	10044
8	31	58	1.43	.37	1.54	2.0	2.29	2.9	.47	10008
48	31	57	1.39	.37	.52	-2.4	.36	-2.3	.85	10048
21	32	58	1.29	.37	1.77	2.8	1.85	2.0	.41	10021
34	32	58	1.29	.37	.67	-1.5	.68	-.9	.78	10034
46	32	57	1.25	.37	.77	-1.0	.71	-.7	.75	10046
18	33	58	1.16	.37	1.60	2.4	1.67	1.6	.44	10018
27	33	58	1.16	.37	1.07	.4	1.07	.3	.63	10027
55	33	57	1.11	.37	.90	-.4	.72	-.7	.73	10055
54	34	57	.98	.37	.65	-1.8	.63	-1.7	.81	10054
10	35	58	.89	.36	1.55	2.4	1.90	1.9	.43	10010
32	35	58	.89	.36	.82	-.8	.62	-.9	.73	10032
58	35	57	.84	.37	.82	-.8	.69	-.7	.75	10058
26	36	58	.75	.36	1.10	.5	.92	.0	.60	10026
36	36	58	.75	.36	.63	-2.2	.39	-1.7	.81	10036
45	36	57	.70	.37	.63	-2.1	.39	-1.6	.81	10045
47	36	57	.70	.37	.61	-2.2	.36	-1.7	.82	10047
50	36	57	.70	.37	.58	-2.5	.35	-1.8	.83	10050
51	36	57	.70	.37	.79	-1.1	.69	-1.2	.75	10051
52	36	57	.70	.37	.62	-2.2	.37	-1.7	.82	10052
25	37	58	.62	.36	1.60	2.8	1.67	1.4	.40	10025
28	37	58	.62	.36	1.44	2.2	3.08	3.1	.46	10028
29	37	58	.62	.36	.81	-1.0	.73	-.5	.73	10029
31	37	58	.62	.36	.67	-2.0	.60	-1.5	.79	10031
43	37	57	.56	.37	.89	-.8	.55	-.9	.71	10043
49	37	57	.56	.37	.82	-1.0	.60	-.8	.75	10049
53	38	57	.43	.37	.70	-1.8	.60	-1.3	.77	10053
30	39	58	.36	.36	.68	-2.1	.60	-1.5	.77	10030
17	40	58	.23	.36	1.22	1.3	1.04	.3	.52	10017
33	40	58	.23	.36	.79	-1.4	.63	-.8	.70	10033
57	40	57	.16	.37	.86	-.9	.57	-.6	.68	10057
11	41	58	.09	.36	1.23	1.4	1.18	.5	.49	10011
41	41	58	.09	.36	1.04	.3	1.07	.3	.58	10041
9	43	58	-.17	.36	1.14	.9	1.64	1.0	.50	10009
14	43	58	-.17	.36	1.38	2.2	1.54	.9	.38	10014
23	43	58	-.17	.36	1.38	2.3	3.94	2.7	.35	10023
35	43	58	-.17	.36	.87	-.8	.50	-.6	.64	10035
42	44	58	-.31	.37	.94	-.4	.52	-.5	.60	10042
56	44	57	-.40	.38	.88	-.7	.47	-.6	.61	10056
16	47	58	-.74	.39	1.16	.9	.79	.0	.43	10016
39	48	58	-.90	.40	1.05	.3	1.55	.8	.40	10039
13	51	58	-1.43	.44	1.04	.2	.59	.0	.40	10013
20	51	58	-1.43	.44	.94	-.2	1.75	.9	.37	10020
37	51	58	-1.43	.44	1.13	.6	1.86	1.0	.30	10037
4	52	58	-1.64	.47	.87	-.4	.44	.0	.44	10004
1	53	58	-1.87	.50	.82	-.5	.32	.0	.45	10001
5	53	58	-1.87	.50	.95	.0	.43	.0	.39	10005
6	53	58	-1.87	.50	.95	-.1	.46	.0	.38	10006
15	53	58	-1.87	.50	1.16	.6	1.27	.7	.22	10015
2	54	58	-2.15	.54	.94	.0	.62	.0	.35	10002
38	54	58	-2.15	.54	.79	-.5	.27	.0	.42	10038
3	55	58	-2.49	.61	.81	-.3	.23	.0	.38	10003
12	55	58	-2.49	.61	1.16	.5	1.12	.8	.16	10012
7	56	58	-2.94	.73	1.02	.2	.69	.0	.22	10007
24	56	58	-2.94	.73	.95	.0	.39	.0	.25	10024

MAP OF ITEMS

```
  5 +
                                                                                                                  S
    |
  4 +
    |
  3 +
    |
    |
  2 +  10019  10022
    |  10040
    |  10044
    |  10008  10048
  1 +  10018  10021  10027  10034  10046  10055                                                                   S
    |  10054
    |  10010  10026  10032  10036  10045  10047  10050  10051  10052  10058
    |  10025  10028  10029  10031  10043  10049
    |  10030  10053
    |  10017  10033  10057
  0 +  10011  10041                                                                                               M
    |  10009  10014  10023  10035
    |  10042  10056
    |
    |  10016
 -1 +  10039
    |                                                                                                             S
    |  10013  10020  10037
    |  10004
    |  10001  10005  10006  10015
 -2 +
    |  10002  10038
    |  10003  10012
    |
    |
 -3 +  10007  10024
```

BEST COPY AVAILABLE

Table 3

Summary of IPARM Analysis of *The Lollipop Test* Items Identified as Potentially Gender Biased

LOLLIPOP 04-01-1993

Item No. 5 Item Name I0005

Logit Item Difficulty -1.87 | Mean Item Score 0.92
Unweighted Total Fit 0.19 | Point Biserial Corr. 0.36
Weighted Total Fit -0.24 | Logit Residual Index 0.05

Distractor/Response Category Analysis Cells = percent of total sample

Ability	Omit	0	1							Mn Abl	N
High		.000	.361							4.12	22
Avg.		.000	.295							1.69	18
Low		.082	.262							-0.47	21
Mean Abl	0.00	-0.66	2.05	0.00	0.00	0.00	0.00	0.00	0.00	1.82	
SD	0.00	0.55	2.04	0.00	0.00	0.00	0.00	0.00	0.00	2.11	
N	0	5	56	0	0	0	0	0	0		61

Total Sample = 61 Total in Item Fit Analysis = 61

Number of omitted persons with unweighted total fit > 99.00 = 0
Number of omitted persons with unweighted total fit < -99.00 = 0

Item No. 5 Item Name I0005

Subpopulation Analysis No. 1 ABILITY Between Fit -0.24

	Ability Groups Low	Middle			High
Predicted Prop.	0.789	0.954	0.997	0.000	0.000
Observed Prop.	0.762	1.000	1.000	0.000	0.000
Difference	-.027	0.046	0.003	0.000	0.000
Raw Score Range	0 - 29	30 - 54	55 - 58	0 - 0	0 - 0
Mean Log Ability	-0.47	1.69	4.12	0.00	0.00
Number of People	21	18	22	0	0

Subpopulation Analysis No. 2 SEX Between Fit 2.19

	MALE	FEMALE			
Predicted Prop.	0.910	0.915	0.000	0.000	0.000
Observed Prop.	0.833	1.000	0.000	0.000	0.000
Difference	-.077	0.085	0.000	0.000	0.000
Mean Log Ability	1.96	1.69	0.00	0.00	0.00
Number of People	30	31	0	0	0

Subpopulation Analysis No. 3 CULTURE Between Fit -0.10

	WHITE	BLACK	HISPANIC		
Predicted Prop.	0.969	0.905	0.808	0.000	0.000
Observed Prop.	0.960	0.958	0.750	0.000	0.000
Difference	-.009	0.053	-.058	0.000	0.000
Mean Log Ability	3.16	.23	0.76	0.00	0.00
Number of People	25	24	12	0	0

LOLLIPOP 04-01-1993

Item No. 28 Item Name I0028

Logit Item Difficulty 0.62 | Mean Item Score 0.64
Unweighted Total Fit 2.63 | Point Biserial Corr. 0.47
Weighted Total Fit 2.15 | Logit Residual Index -0.97

Distractor/Response Category Analysis Cells = percent of total sample

Ability	Omit	0	1							Mn Abl	N
High		.049	.311							4.12	22
Avg.		.098	.197							1.69	18
Low		.213	.131							-0.47	21
Mean Abl	0.00	0.52	2.57	0.00	0.00	0.00	0.00	0.00	0.00	1.82	
SD	0.00	1.78	1.89	0.00	0.00	0.00	0.00	0.00	0.00	2.11	
N	0	22	39	0	0	0	0	0	0		61

Total Sample = 61 Total in Item Fit Analysis = 61

Number of omitted persons with unweighted total fit > 99.00 = 0
Number of omitted persons with unweighted total fit < -99.00 = 0

Item No. 28 Item Name I0028

Subpopulation Analysis No. 1 ABILITY Between Fit 2.14

	Ability Groups Low	Middle			High
Predicted Prop.	0.265	0.691	0.965	0.000	0.000
Observed Prop.	0.381	0.667	0.864	0.000	0.000
Difference	0.116	-.025	-.101	0.000	0.000
Raw Score Range	0 - 29	30 - 54	55 - 58	0 - 0	0 - 0
Mean Log Ability	-0.47	1.69	4.12	0.00	0.00
Number of People	21	18	22	0	0

Subpopulation Analysis No. 2 SEX Between Fit 2.27

	MALE	FEMALE			
Predicted Prop.	0.653	0.634	0.000	0.000	0.000
Observed Prop.	0.533	0.742	0.000	0.000	0.000
Difference	-.120	0.108	0.000	0.000	0.000
Mean Log Ability	1.96	1.69	0.00	0.00	0.00
Number of People	30	31	0	0	0

Subpopulation Analysis No. 3 CULTURE Between Fit -0.99

	WHITE	BLACK	HISPANIC		
Predicted Prop.	0.631	0.568	0.402	0.000	0.000
Observed Prop.	0.800	0.583	0.417	0.000	0.000
Difference	.031	0.016	0.014	0.000	0.000
Mean Log Ability	3.16	1.23	0.26	0.00	0.00
Number of People	25	24	12	0	0

BEST COPY AVAILABLE

Table 4

Summary of IPARM Analysis of *The Lollipop Test* Items Identified as Potentially Culture Biased

LOLLIPOP 04-01-1993

Item No. 9 Item Name I0009

Logit Item Difficulty	-0.17	Mean Item Score	0.75
Unweighted Total Fit	0.87	Point Biserial Corr.	0.49
Weighted Total Fit	0.95	Logit Residual Index	-0.17

Distractor/Response Category Analysis Cells = percent of total sample

Ability	Omit	0	1							Mn Abl	N
High		.016	.344							4.12	22
Avg.		.082	.213							1.69	18
Low		.148	.197							-0.47	21
Mean Abl	0.00	0.02	2.41	0.00	0.00	0.00	0.00	0.00	0.00	1.82	
SD	0.00	1.24	1.98	0.00	0.00	0.00	0.00	0.00	0.00	2.11	
N	0	15	46	0	0	0	0	0	0		61

Total Sample = 61 Total in Item Fit Analysis = 61

Number of omitted persons with unweighted total fit > 99.00 = 0
Number of omitted persons with unweighted total fit < -99.00 = 0

LOLLIPOP 04-01-1993

Item No. 19 Item Name I0019

Logit Item Difficulty	2.25	Mean Item Score	0.44
Unweighted Total Fit	2.79	Point Biserial Corr.	0.47
Weighted Total Fit	2.74	Logit Residual Index	-0.86

Distractor/Response Category Analysis Cells = percent of total sample

Ability	Omit	0	1							Mn Abl	N
High		.098	.262							4.12	22
Avg.		.180	.115							1.69	18
Low		.279	.066							-0.47	21
Mean Abl	0.00	0.95	2.93	0.00	0.00	0.00	0.00	0.00	0.00	-1.82	
SD	0.00	1.69	2.04	0.00	0.00	0.00	0.00	0.00	0.00	2.11	
N	0	34	27	0	0	0	0	0	0		61

Total Sample = 61 Total in Item Fit Analysis = 61

Number of omitted persons with unweighted total fit > 99.00 = 0
Number of omitted persons with unweighted total fit < -99.00 = 0

LOLLIPOP 04-01-1993

Item No. 25 Item Name I0025

Logit Item Difficulty	0.62	Mean Item Score	0.64
Unweighted Total Fit	1.16	Point Biserial Corr.	0.47
Weighted Total Fit	2.79	Logit Residual Index	-0.20

Distractor/Response Category Analysis Cells = percent of total sample

Ability	Omit	0	1							Mn Abl	N
High		.016	.344							4.12	22
Avg.		.164	.131							1.69	18
Low		.180	.164							-0.47	21
Mean Abl	0.00	0.51	2.56	0.00	0.00	0.00	0.00	0.00	0.00	1.82	
SD	0.00	1.66	2.04	0.00	0.00	0.00	0.00	0.00	0.00	2.11	
N	0	22	39	0	0	0	0	0	0		61

Total Sample = 61 Total in Item Fit Analysis = 61

Number of omitted persons with unweighted total fit > 99.00 = 0
Number of omitted persons with unweighted total fit < -99.00 = 0

Item No. 9 Item Name I0009

Subpopulation Analysis No. 1 ABILITY Between Fit 1.12

	Ability Groups Low	Middle	High		
Predicted Prop.	0.432	0.812	0.984	0.000	0.000
Observed Prop.	0.571	0.722	0.955	0.000	0.000
Difference	0.139	-.089	-.029	0.000	0.000
Raw Score Range	0 - 29	30 - 54	55 - 58	0 - 0	0 - 0
Mean Log Ability	-0.47	1.69	4.12	0.00	0.00
Number of People	21	18	22	0	0

Subpopulation Analysis No. 2 SEX Between Fit 0.44

	MALE	FEMALE			
Predicted Prop.	0.746	0.740	0.000	0.000	0.000
Observed Prop.	0.800	0.710	0.000	0.000	0.000
Difference	0.054	-.030	0.000	0.000	0.000
Mean Log Ability	1.96	1.69	0.00	0.00	0.00
Number of People	30	31	0	0	0

Subpopulation Analysis No. 3 CULTURE Between Fit 2.28

	WHITE	BLACK	HISPANIC		
Predicted Prop.	0.891	0.697	0.527	0.000	0.000
Observed Prop.	0.960	0.542	0.750	0.000	0.000
Difference	0.069	-.156	0.223	0.000	0.000
Mean Log Ability	3.14	1.23	0.26	0.00	0.00
Number of People	25	24	12	0	0

Item No. 19 Item Name I0019

Subpopulation Analysis No. 1 ABILITY Between Fit 1.98

	Ability Groups Low	Middle	High		
Predicted Prop.	0.069	0.397	0.867	0.000	0.000
Observed Prop.	0.190	0.389	0.727	0.000	0.000
Difference	0.122	-.008	-.139	0.000	0.000
Raw Score Range	0 - 29	30 - 54	55 - 58	0 - 0	0 - 0
Mean Log Ability	-0.47	1.69	4.12	0.00	0.00
Number of People	21	18	22	0	0

Subpopulation Analysis No. 2 SEX Between Fit -1.20

	MALE	FEMALE			
Predicted Prop.	0.448	0.425	0.000	0.000	0.000
Observed Prop.	0.467	0.419	0.000	0.000	0.000
Difference	-.001	-.006	0.000	0.000	0.000
Mean Log Ability	1.96	1.69	0.00	0.00	0.00
Number of People	30	31	0	0	0

Subpopulation Analysis No. 3 CULTURE Between Fit 2.93

	WHITE	BLACK	HISPANIC		
Predicted Prop.	0.644	0.335	0.216	0.000	0.000
Observed Prop.	0.560	0.292	0.500	0.000	0.000
Difference	-.106	-.043	0.284	0.000	0.000
Mean Log Ability	3.14	1.23	0.26	0.00	0.00
Number of People	25	24	12	0	0

Item No. 25 Item Name I0025

Subpopulation Analysis No. 1 ABILITY Between Fit 2.72

	Ability Groups Low	Middle	High		
Predicted Prop.	0.265	0.691	0.965	0.000	0.000
Observed Prop.	0.476	0.444	0.955	0.000	0.000
Difference	0.211	-.247	-.010	0.000	0.000
Raw Score Range	0 - 29	30 - 54	55 - 58	0 - 0	0 - 0
Mean Log Ability	-0.47	1.69	4.12	0.00	0.00
Number of People	21	18	22	0	0

Subpopulation Analysis No. 2 SEX Between Fit 0.62

	MALE	FEMALE			
Predicted Prop.	0.653	0.634	0.000	0.000	0.000
Observed Prop.	0.700	0.581	0.000	0.000	0.000
Difference	0.047	-.053	0.000	0.000	0.000
Mean Log Ability	1.96	1.69	0.00	0.00	0.00
Number of People	30	31	0	0	0

Subpopulation Analysis No. 3 CULTURE Between Fit 2.08

	WHITE	BLACK	HISPANIC		
Predicted Prop.	0.831	0.568	0.402	0.000	0.000
Observed Prop.	0.800	0.458	0.667	0.000	0.000
Difference	-.031	-.109	0.264	0.000	0.000
Mean Log Ability	3.14	1.23	0.26	0.00	0.00
Number of People	25	24	12	0	0

BEST COPY AVAILABLE

TABLE 5

Summary of Fit Statistics for Selected Items

Items suggested by calibration:

	Total Unweighted Fit	Between Fit Sex	Between Fit Culture
8	2.41	-0.24	1.52
10	1.59	-0.70	0.21
14	0.77	-0.19	0.43
18	1.32	-0.36	1.11
19	2.79	-1.20	2.93
21	1.75	-1.05	0.39
22	2.28	0.02	0.53
23	2.34	-0.90	-0.13
25	1.14	0.62	2.08
28	2.63	2.27	-0.99

Items suggested by between fit analysis:

Gender

5	0.19	2.19	-0.10
28	2.63	2.27	-0.99

Culture

9	0.87	0.44	2.28
19	2.79	-1.20	2.93
25	1.14	0.62	2.08

Seventh International Objective Measurement Workshop

Updated Information Sheet

Time: Saturday, April 10 to Sunday, April 11

Place: 206 White Hall
Emory University
Atlanta, GA 30322
USA

Co-Chairs: George Engelhard, Jr. , Emory University and Judy Monsaas, West Georgia College

Phone: George Engelhard (404-727-0607 at office and 404-525-1115 at home)

Saturday Night Dinner: Dinner will be at Jagger's Restaurant (Number 121 on your map) at 6:00 p.m.

Van Schedule: A van will leave from Emory Inn to White Hall at 8:00 and 8:15 a.m. on Saturday and Sunday.

Schedule Changes:

Pender Pedler will not present a paper during Session 6 (11:00-12:00)

Ben Wright will present a paper entitled "The significance of divisibility in Rasch measurement" during Session 5 (8:30-10:30) in place of the last two presentations ("Facets as ANOVA" and "Measuring with unexpected relevant obversations")

[Map on back]

USF POSITION VACANCY LISTING

Division of Personnel Services
4202 E. Fowler Avenue
Tampa, Florida 33620-7800
Phone (813) 974-2974 TTY# (813) 974-2218
Posted 09/16/94

Title/Department	Position #	Biweekly Salary Range	Deadline Date	Contact Name Building Location Telephone #	Qualifications
Senior Accountant Finance and Accounting	70011	$840.99 - $1,463.66	09/22/94	Ginger Pickern ADM 147 974-6062	Bachelor's degree in Business, Accounting or Finance; and three years of professional accounting experience. Appropriate college coursework may substitute at an equivalent rate for the required experience. **PREFER** experience with Lotus 1-2-3 and State Accounting System (SAMAS).
Senior Clerk Finance and Accounting	77044	$527.93 - $894.26	09/22/94		High school diploma and two years of clerical experience. Appropriate college coursework or vocational/technical training may substitute at an equivalent rate for the required experience. **PREFER** CRT (terminal or personal computer), inventory experience, and ability to move 70 lbs. Valid Florida driver's license required.
Senior Environmental Health and Safety Specialist Environmental Health and Safety	79879	$866.79 - $1,534.04	09/22/94	Personnel Services SVC 2172 974-2974	Bachelor's degree in an appropriate area of specialization and two years of appropriate experience. Appropriate college coursework may substitute at an equivalent rate for the required experience. **PREFER** bachelor's degree in a science related field with two years experience in laboratory and chemical safety, environmental sanitation and water quality.

To apply for positions listed, please refer to the [illegible] column. You [illegible] application to Personnel or take your [illegible]

BEST COPY AVAILABLE

USF POSITION VACANCY LISTING

Division of Personnel Services
4202 E. Fowler Avenue
Tampa, Florida 33620-7800
Phone (813) 974-2974 TTY# (813) 974-2218
Posted 09/16/94

Title/Department	Position #	Biweekly Salary Range	Deadline Date	Contact Name Building Location Telephone #	Qualifications
Senior Fiscal Assistant Finance and Accounting	77040	$554.41 - $944.94	09/22/94	Ginger Pickern ADM 147 974-6062	High school diploma and three years of clerical/accounting experience. Appropriate college coursework or vocational/technical training may substitute at an equivalent rate for the required experience. **PREFER** terminal/computer, SAMAS, and accounts receivable experience.
Senior Library Technical Assistant Center for Urban Transportation Research Engineering **Time-Limited**	67315	$584.02 - $1,000.24	09/27/94	Cindy Wooten ENB 118 974-3120	High school diploma and four years of appropriate experience. Appropriate college coursework or vocational/technical training may substitute at an equivalent rate for the required experience. **PREFER** Bachelor's degree in Civil (transportation) Engineering or related field, work experience with ProCite, project budget management, WordPerfect with windows, database software and mainframe operation. Desire excellent verbal and written communication skills, and proven initiative in job assignments.
Word Processing Operator Institute on Black Life Academic Affairs	73863	$527.93 - $894.26	09/27/94	Personnel Services SVC 2172 974-2974	High school diploma and two years of office experience. Appropriate college coursework or vocational/technical training may substitute at an equivalent rate for the required experience. Attainment of a score of at least 35 cwpm on an approved typing test. **PREFER** personal computer, word processing, and methods of data collection experience. Desire proficiency in using WordPerfect 6.0, Pagemaker, accounting software (Lotus or QuadroPro), INTERNET (e-mail and gopher), and an undergraduate degree.

To apply for [illegible] USF USPS
[illegible] application [illegible] to the hiring department.